Common Factor Analysis Versus Principal Component Analysis: Choice for Symptom Cluster Research

*Hee-Ju Kim**, PhD, RN

Full-time Instructor, Department of Nursing, University of Ulsan, Ulsan, South Korea

Purpose The purpose of this paper is to examine differences between two factor analytical methods and their relevance for symptom cluster research: common factor analysis (CFA) versus principal component analysis (PCA).
Methods Literature was critically reviewed to elucidate the differences between CFA and PCA. A secondary analysis ($N = 84$) was utilized to show the actual result differences from the two methods.
Results CFA analyzes only the reliable common variance of data, while PCA analyzes all the variance of data. An underlying hypothetical process or construct is involved in CFA but not in PCA. PCA tends to increase factor loadings especially in a study with a small number of variables and/or low estimated communality. Thus, PCA is not appropriate for examining the structure of data.
Conclusion If the study purpose is to explain correlations among variables and to examine the structure of the data (this is usual for most cases in symptom cluster research), CFA provides a more accurate result. If the purpose of a study is to summarize data with a smaller number of variables, PCA is the choice. PCA can also be used as an initial step in CFA because it provides information regarding the maximum number and nature of factors. In using factor analysis for symptom cluster research, several issues need to be considered, including subjectivity of solution, sample size, symptom selection, and level of measure. [*Asian Nursing Research* 2008;2(1):17–24]

Key Words cancer, nursing, statistics, symptom clusters

INTRODUCTION

Symptom clusters are stable groups of simultaneously occurring and interrelated symptoms (Dodd, Miaskowski, & Paul, 2001; Kim, McGuire, Tulman, & Barsevick, 2005). The attention given to symptom cluster research in oncology has increased enormously since its first introduction in 2001 by Dodd and her colleagues (Barsevick, Whitmer, Nail, Beck, & Dudley, 2006; Miaskowski, Dodd, & Lee, 2004). In fact, the largest oncology nursing group in the United States, the Oncology Nursing Society (2007), highlights this research field as a top priority, because it is promising to develop more efficient symptom assessment and management strategies from symptom cluster research. Although symptoms in

***Correspondence to**: Hee-Ju Kim, PhD, RN, Full-time Instructor, Department of Nursing, University of Ulsan, Nan-gu Dea-Hak-Ro 102, Ulsan 680-749, Korea.
E-mail: heeju06@gmail.com

cancer patients have a tendency to occur together, the traditional symptom assessment or management strategies have mainly focused on an individual symptom (e.g., pain) (Dodd et al.). Symptom cluster researchers in oncology cast a doubt on the traditional single-symptom approach. Thus, researchers have tried to identify a group of symptoms that can be assessed or managed together, that is, a symptom cluster. This type of approach to multiple symptoms will be more cost effective than the single-symptom approach (Kim et al.).

This paper builds upon this author's previous work: concept analysis of symptom clusters (Kim et al., 2005). The concept analysis paper examined how the concept of symptom clusters were defined and utilized in diverse disciplines (psychology/psychiatry and some areas of medicine, and nursing) by critically reviewing literature. The literature review process also indicated that various statistical methodologies have been used in diverse disciplines to identify symptom clusters: correlation and related measures of association, graphical modeling, factor analysis, and cluster analysis (Kim & Abraham, in press). Among these methods, factor analysis was the most commonly used methodology (Kim & Abraham, in press; see bibliography for article exemplars using factor analysis). In reviewing literature for the concept analysis paper (Kim et al.), it was also found that two types of factor analysis have been used: common factor analysis (CFA) and principal component analysis (PCA). Many researchers have not explained why they chose one method over the other. However, these two methods are quite different in terms of their purposes, mathematical backgrounds, and results. The purpose of this paper is to discuss the ways these methods differ and circumstances in symptom cluster research where each method is preferred.

ANALYZING SYMPTOM CLUSTERS USING FACTOR ANALYSIS

Factor analysis is a statistical tool used to account for observed correlations among many variables, particularly when "causation is complex and multivariate and basic concepts have been elusive" (Cattell, 1988, p. 131). The invention and development of factor analytical techniques lies in the history of instrument development in psychology, such as personality and intelligence tests. Still, the most common use of this technique is for test or measurement development. However, factor analysis can also be used to identify symptom clusters because it shows a group of symptoms that are highly associated with each other (Kim & Abraham, in press). In particular, factor analysis has a unique strength in identifying symptom clusters induced by a common underlying process or basis (i.e., common factors) because of its mathematical assumption discussed in the following section.

FACTOR ANALYSIS

The basic concept of factor analysis

Factor analysis assumes that a common variable (C) is responsible for the creation of two variables, A and B, and thus it is also responsible for the observed correlations/associations between variables. Such a common variable (C) is called a *common factor* in factor analysis. It is often named as a latent variable or underlying factor because such a variable is not directly measured (Gorsuch, 1983; Kim & Mueller, 1978b).

In this context, factor analysis can be described as a statistical technique used for finding common factors which explain the correlation among variables. Factor analysis also shows a group of variables which are highly associated with, and thus are representing, a common factor. Through this process, the structures, dimensions, or underlying processes of the data are also identified. Figure 1 illustrates what the structures, dimensions, or underlying processes of data detected by factor analysis are.

In the analysis process, identifying a common factor starts from the correlation matrix of variables. The first common factor will be determined when it explains the most correlations in this matrix. Again, the common factor is not measured but is generated. Providing meaning to the common factor is a theoretical procedure rather than a statistical one.

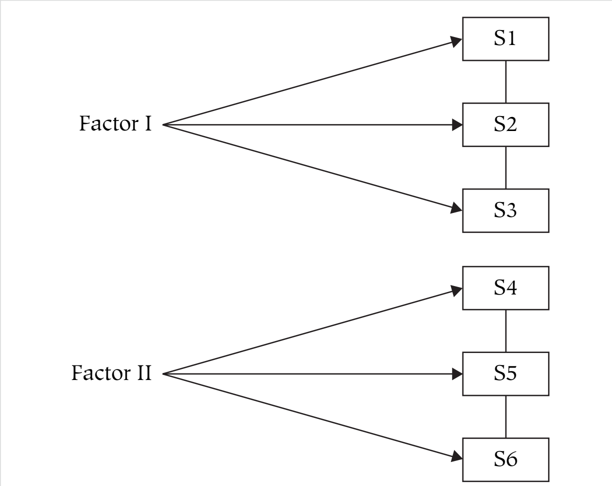

Figure 1. Underlying dimension, process, and structure of the data. *Note*. S represents an observed variable. Factors, here, are common factors. Arrows indicate that factors create variables. A line indicates that there is a correlation between variables. This diagram illustrates relationships between factors and observed variables, and how factor analysis can show the underlying process, dimensions, and structure of the data (observed variables).

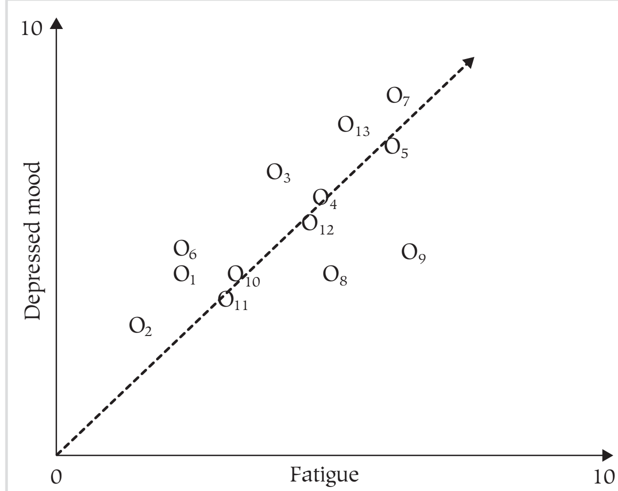

Figure 2. Visual explanation for factor analysis. *Note*. This is the simplest example with only two symptoms (fatigue and depressed mood). Each patient's score on the severity of the two symptoms is marked in the two dimensional space (O). Dotted arrow indicates a common factor. The common factor is inserted where the covariance of the two symptoms is best explained.

Graphical explanation might be helpful to better understand the factor analysis procedure. Let's imagine a two-dimensional space with the X axis indicating fatigue and the Y axis indicating depressed mood. Each patient's score on the severity of the two symptoms (fatigue and depressed mood) can be marked in the two-dimensional space (see Figure 2). Then, a factor is inserted into this two-dimensional variable space. The position of a factor, in the case of principal axis factoring, is determined by the sum of the squared deviations (SSD) of all the data points from a factor; where SSDs are at a minimum, the factor will be inserted (Kim & Mueller, 1978a; Tryon & Bailey, 1970). That is, a factor is generated to explain correlations between variables.

CFA versus PCA

Now, what are the differences between the two types of factor analysis, CFA versus PCA? Note that precisely speaking, what we are searching for in PCA is not a *factor*, but a *component*. However, the term *factor* has been used to refer to both components and factors in the literature and here as well. Whenever the distinction is critical, a specific term is used in this paper.

Examining the equations used in each method (note: equations are adapted from Gorsuch, 1983, p. 51, 53) will be helpful to understand the differences between the two factor analysis methods. The relationships between observed variables and factors can be described in a regression equation, considering the mathematical assumptions behind factor analysis. PCA can be illustrated by the equation: $Z = FP$, where Z is "the standardized score data matrix" of original variables, F is "the standardized factor score matrix", and P is "the factor by variable weight matrix (factor × variable weight matrix)" (Gorsuch). Simply, we can assume that Z represents the variance of a variable, F represents a variance of common factors (or, a *common variance* of a Z variable with other variables in analysis), and P represents a coefficient showing how F and Z are related. Note that common variance (or covariance) can be considered as types of correlation for simpler understanding.

The CFA equation is: $Z = FP + UD$

The difference between the two equations is the last component (i.e., UD). Simply speaking, UD represents a unique variance of a variable. UD contains both the *unreliable variance* of measurement error and the *reliable variance* which does not overlap with common variance (Fabrigar, Wegener, MacCallum, & Strahan, 1999). Figure 3 illustrates common variance and unique variance. The areas of A and B in Figure 3 represent the unique variance of variables Z_1 and Z_2, respectively.

PCA assumes that the common variance (C in Figure 3) becomes maximized and there is no unique variance (A and B) in each variable. Whereas CFA assumes that there is a substantial amount of unique variance as well as reliable common variance.

From the two equations, in PCA, all the variance of a variable (Z) comes from common factors (F) (strictly speaking components); in other words, if we know the factor (factor scores) and their weights, we can recalculate the original variables. Also, a matrix of correlation between variables can be recalculated from weights and correlation among factors. However, in CFA, without knowing or estimating the unique variance of variables (UD), neither variables nor the correlation between them can be exactly recalculated. In analysis, CFA differentiates unique variance from common variance, by estimating unique variance and excluding it from the analysis (Tabachnick & Fidell, 2001).

Another difference between the two procedures is the matrix from which factors are extracted. The diagonal of the matrix is the unity in PCA (i.e., 1) while it is the estimated communality (i.e., common variance) in CFA. In other words, PCA analyzes all of the variance of the set of variables (common variance and unique variance), but CFA analyzes only the common variance (covariance or correlation) of the set of variables.

Consequently, the purpose of PCA is not to explain the correlations among variables, but to explain as much variance as possible in the data. In the analysis procedure of PCA, the existence of hypothetical underlying factors is not necessary, and the component is simply a combination of correlated variables (Kim & Mueller, 1978a; Tabachnick & Fidell, 2001). In this context, the main purpose of PCA is to summarize many variables into a smaller number of components: that is, data reduction. On the other hand, CFA finds a factor model that would best reproduce the observed correlation, and thus it is aimed at explaining the correlation between variables (Kim & Mueller; Lorr, 1983). The difference between *component* and *factor* can be stated as the following: variables create components, while factors create variables (Tabachnick & Fidell). This notion is illustrated in Figure 4.

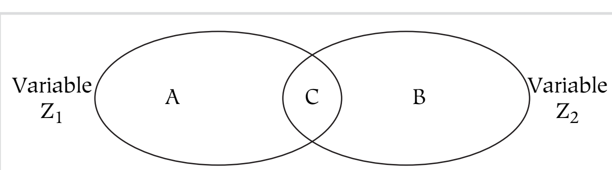

Figure 3. Common variance versus unique variance. *Note*. Area C indicates the common variance of variable Z_1 with variable Z_2. Areas A and B indicate the unique variance of variables Z_1 and Z_2, respectively. The unique variance includes the unreliable variance of measurement error and reliable variance which does not overlap with common variance.

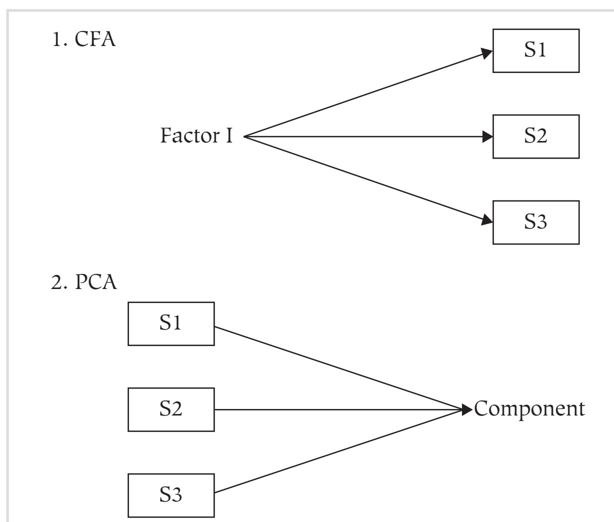

Figure 4. CFA versus PCA. *Note*. The directions of arrows are different in CFA and PCA.

SELECTING FACTOR ANALYSIS FOR SYMPTOM CLUSTER RESEARCH

The above theoretical differences between the two methods (CFA and PCA) will have practical implications on research only when the findings are different. There have been discussions regarding pragmatic differences between the two methods. Many researchers agree that CFA and PCA produce similar solutions (i.e., results) in cases with a large number of variables (e.g., 30 or more) and/or high estimated communality (simply speaking, common variance of variables) (Gorsuch, 1983). In other contexts, however, the results from the two are different. In most cases of symptom cluster research, researchers are more likely to have different results from PCA and CFA. This is because it is very rare to include more than 30 different symptoms in analyses for symptom cluster research, unless redundant symptoms are included. Also, correlations (or common variance) among clinical symptoms are not very high. Because numerous factors affect symptom experience in cancer patients (e.g., drugs used for cancer treatments or for symptom management, stage of disease, comorbid conditions, etc.), the pattern of associations between symptoms may not be very consistent across subjects, and consequently correlations (r) are usually not that high.

Table 1 presents findings from a small simulation study showing the differences in finding from the two methods (PCA and CFA). In this simulation study ($N=84$), each method was applied to the same data. One factor was extracted out of eight variables. The eight variables comprise a subscale of a measure. As shown in Table 1, factor loadings in PCA were higher than in CFA.

Table 1

Differences Between PCA and CFA in Factor Loadings from a Simulation Study (N = 84)

No. of variables	CFA factor loadings	PCA factor loadings
1	.41	.49
2	.61	.67
3	.50	.58
4	.65	.70
5	.53	.61
6	.54	.63
7	.70	.75
8	.38	.46

Note. This simulation study is a secondary data analysis. Data were collected from Korean college students. For this simulation study, the eight variables comprising one subscale (Perceived Barriers on Exercise) of a measure (the Decision Balance, developed by Kim, 2002) were taken and were factor analyzed.

Then, which method is more appropriate for symptom cluster research? CFA sounds more attractive. There are several reasons for this. First of all, there is no variable (in this case, a symptom) that is error free, and all of whose variance can be produced by common factors (Gorsuch, 1983). Note that PCA assumes this. We do not expect the variance of a variable to be totally the function of factors. Second, PCA is not appropriate to examine the structure of the variables. There have been consistent reports that PCA inflates factor loadings, consequently altering the factor structure, particularly in cases with a small number of variables (Fabrigar et al., 1999; Gorsuch; Snook & Gorsuch, 1989). In addition, the main purposes of the two techniques are different: PCA for data reduction versus CFA for explanation of the correlations among observed variables with hypothetical variables. If the purpose of the study is data reduction, PCA is the choice (Tabachnick & Fidell, 2001). However, if the purpose of the study is to explain the correlations among variables and to examine the structure of the data, CFA is favored (Fabrigar et al.; Tabachnick & Fidell). The latter seems to be the case in symptom cluster research. Nonetheless, PCA can also be used as an initial step in CFA because it can provide information regarding the maximum number and nature of factors (Tabachnick & Fidell).

PRACTICAL ISSUES IN USING FACTOR ANALYSIS FOR SYMPTOM CLUSTER RESEARCH

Several practical issues need to be considered in using factor analysis (including PCA) in symptom cluster research. First, the best factor model (or result) may be subjective. Although guidelines for decision making involved in the analysis process have been developed, each analyst can make a slightly different decision for the same data and thereby have a different result. Therefore, establishing strategies to find robust and reliable solutions (i.e., factor models) is essential. It is also necessary to report the criteria used to find the best solution. Of course, those criteria should be theoretically and practically sound.

Second, establishing a large enough sample size can be a challenge because patients with symptoms are suffering from cancer or its treatments and they are less likely to participate in a study. There is no clear standard to calculate sample size for factor analysis. Power analysis is not relevant for factor analysis because no inferential test is involved. The most widely used principle for the sample size calculation is at least 10 subjects per variable. More conservative statistical experts recommend a sample size of a minimum of 300 subjects for a reliable solution (Tabachnick & Fidell, 2001). Practically, it is possible to conduct factor analysis with a smaller sample (e.g., $N = 100$). Factor analysis results with 100 cases are often found in the literature. However, the results should be carefully interpreted with regard to sample size. If the sample size is not big enough, the results may not be reliable. That is, the results may be different from one sample to another, and the results from a sample may not reveal the true findings in the population. When working with a small sample, researchers may want to provide other evidence supporting the reliability of their findings.

Third, factor analysis causes technical difficulties in analyzing nominal level variables (e.g., presence or absence of a symptom) and therefore variables need to be measured at a level higher than nominal level.

Fourth, the selection of symptoms which will be included in analysis is an important issue for factor analysis, especially when the purpose of the study is to identify underlying structures (dimensions) of data. This is because the common factors can be over- or under-identified in analysis when the selected symptoms over- or under-represent the domain of interest in the variable selection (Fabrigar et al., 1999). For example, Gift, Jablonski, Stommel, and Given (2004) conducted factor analysis on symptoms to identify symptom clusters in lung cancer patients and found a symptom cluster (fatigue, nausea, weakness, appetite loss, weight loss, altered taste, vomiting). It appears that their analysis did not include depressed mood. It is a possible scenario that one common factor of symptoms which depressed mood and other associated symptoms represent might not have been identified because of the absence of depressed mood. Indeed, some symptoms which seem to be related to depressed mood were included in their study, such as lack of sexual interest, difficulty concentrating, and/or trouble sleeping, but these symptoms did not form a symptom cluster representing a common factor.

CONCLUSION

Factor analysis is the most popular statistical procedure to identify symptom clusters in other disciplines (e.g., general medicine, psychology, psychiatry), and its popularity is growing in oncology symptom cluster research (Chen & Tseng, 2006; Gift et al., 2004). Factor analysis may identify groups of symptoms interrelated due to a common underlying source (factor) and it can be a good start for examination of a common cause of symptoms (Kim & Abraham, in press). This paper underscores that researchers need to be aware of the fact that there are variants of factor analysis and each variant can yield different findings in some circumstances. Only the proper use of a statistical method yields sound findings. Researchers need to be careful in selecting a statistical method. It is a dangerous approach to choose a methodology simply because a certain method was used for a similar research purpose in previous studies. Researchers need to establish extensive knowledge regarding a methodology before choosing one method over the other. Close cooperation with statistical experts can also be very helpful in understanding the application of a statistical method as well as in interpreting results.

References

Barsevick, A. M., Whitmer, K., Nail, L. M., Beck, S. L., & Dudley, W. N. (2006). Symptom cluster research: Conceptual, design, measurement, and analysis issues. *Journal of Pain and Symptom Management*, *31*, 85–95.

Cattell, R. B. (1988). The meaning and strategic use of factor analysis. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 131–201). New York: Plenum Press.

Chen, M. L., & Tseng, H. C. (2006). Symptom clusters in cancer patients. *Supportive Care in Cancer, 14*, 825–830.

Dodd, M. J., Miaskowski, C., & Paul, S. M. (2001). Symptom clusters and their effect on the functional status of patients with cancer. *Oncology Nursing Forum*, *28*, 465–470.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272–299.

Gift, A. U., Jablonski, A., Stommel, M., & Given, W. (2004). Symptom clusters in elderly patients with lung cancer. *Oncology Nursing Forum*, *31*, 203–210.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kim, H.-J., & Abraham, I. (in press). Statistical approaches to modeling symptom clusters in cancer patients. *Cancer Nursing*.

Kim, H.-J., McGuire, D. B., Tulman, L., & Barsevick, A. M. (2005). Symptom clusters: Concept analysis and clinical implications for cancer nursing. *Cancer Nursing*, *28*, 270–282.

Kim, J., & Mueller, C. W. (1978a). *Factor analysis: Statistical methods and practical issues*. Beverly Hills, CA: Sage Publications.

Kim, J., & Mueller, C. W. (1978b). *Introduction to factor analysis: What it is and how to do it*. Beverly Hills, CA: Sage Publications.

Kim, Y. H. (2002). Adolescents' stage of change, decision balance, self-efficacy in exercise: Application of transtheoretical model. *Korean Journal of Psychology*, *13*, 1–19.

Lorr, M. (1983). *Cluster analysis for social scientists*. San Francisco: Jossey-Bass.

Miaskowski, C., Dodd, M., & Lee, K. (2004). Symptom clusters: The new frontier in symptom management research. *Journal of the National Cancer Institute Monologues*, *32*, 17–21.

Oncology Nursing Society. (2007). *Oncology nursing society 2005–2009 research agenda*. Retrieved August 28, 2007, from http://www.ons.org/research/information/documents/pdfs/ResearchAgendaExecutiveSummary10-24-07.pdf

Snook, S. C., & Gorsuch R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin*, *106*, 148–154.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). New York: HarperCollins College Publisher.

Tryon, R. C., & Bailey, D. E. (1970). *Cluster analysis*. New York: McGraw-Hill.

Bibliography

Listed articles include those reviewed for this author's previous work on concept analysis (Kim et al., 2005), in which a total of 78 articles were examined. Because random sampling from the database was involved in selecting those articles, the articles fairly well represent the methodology used for symptom cluster research in diverse disciplines. Only those which clearly indicated factor analysis as a method ($n = 26$) are listed here, and articles recently published in oncology are also listed and marked with an asterisk (*).

Abraham, G., Nair, C., Tracy, J. I., Simpson, G. M., & Josiassen, R. C. (1997). The effects of clozapine on symptom clusters in treatment-refractory patients. *Journal of Clinical Psychopharmacology*, *17*, 49–53.

Alvir, J. M., & Thys-Jacobs, S. (1991). Premenstrual and menstrual symptom clusters and response to calcium treatment. *Psychopharmacology Bulletin*, *27*, 145–148.

Amdur, R. L., & Liberzon, I. (2001). The structure of posttraumatic stress disorder symptoms in combat veterans: A confirmatory factor analysis of the impact of event scale. *Journal of Anxiety Disorders, 15*, 345–357.

Asmundson, G. J., Frombach, I., McQuaid, J., Pedrelli, P., Lenox, R., & Stein, M. B. (2000). Dimensionality of posttraumatic stress symptoms: a confirmatory factor analysis of DSM-IV symptom clusters and other symptom models. *Behaviour Research and Therapy*, *38*, 203–214.

Barrett, B., Locken, K., Maberry, R., Schwamman, J., Brown, R., Bobula, J., et al. (2002). The Wisconsin Upper Respiratory Symptom Survey (WURSS): A new research instrument for assessing the common cold. *Journal of Family Practice*, *51*, 265.

Bassett, A. S., Bury, A., & Honer, W. G. (1994). Testing Liddle's three-syndrome model in families with schizophrenia. *Schizophrenia Research*, *12*, 213–221.

Calvaresi, E., & Bryan, J. (2003). Symptom experience in Australian men and women in midlife. *Maturitas*, *44*, 225–236.

*Chen, M. L., & Tseng, H. C. (2006). Symptom clusters in cancer patients. *Supportive Care in Cancer*, *14*, 825–830.

Clark, L. A. (1990). Toward a consensual set of symptom clusters for assessment of personality disorder. In J. Butcher & C. Spielberger (Eds.), *Advances in personality assessment* (Vol. 8, pp. 243–266). Hillsdale, NJ: Lawrence Erlbaum Associates.

de Groot, C. M., Bornstein, R. A., & Baker, G. B. (1995). Obsessive-compulsive symptom clusters and urinary amine correlates in Tourette syndrome. *Journal of Nervous and Mental Disease*, *183*, 224–230.

Dunn, R. T., Kimbrell, T. A., Ketter, T. A., Frye, M. A., Willis, M. W., Luckenbaugh, D. A., et al. (2002). Principal components of the Beck Depression Inventory and regional cerebral metabolism in unipolar and bipolar depression. *Biological Psychiatry*, *51*, 387–399.

*Gift, A. G., Jablonski, A., Stommel, M., & Given, W. (2004). Symptom clusters in elderly patients with lung cancer. *Oncology Nursing Forum*, *31*, 203–210.

*Gift, A. G., Stommel, M., Jablonski, A., & Given, W. (2003). A cluster of symptoms over time in patients with lung cancer. *Nursing Research*, *52*, 393–400.

Gulick, E. E. (1989). Model confirmation of the MS-related symptom checklist. *Nursing Research*, *38*, 147–153.

Ho, S. C., Chan, S. G., Yip, Y. B., Cheng, A., Yi, Q., & Chan, C. (1999). Menopausal symptoms and symptom clustering in Chinese women. *Maturitas*, *33*, 219–227.

Holzemer, W. L., Henry, S. B., Nokes, K. M., Corless, I. B., Brown, M. A., Powell-Cope, G. M., et al. (1999). Validation of the Sign and Symptom Check-list for persons with HIV disease (SSC-HIV). *Journal of Advanced Nursing*, *30*, 1041–1049.

Hunter, M., Battersby, R., & Whitehead, M. (1986). Relationships between psychological symptoms, somatic complaints and menopausal status. *Maturitas*, *8*, 217–228.

Mitchell, E. S., & Woods, N. F. (1996). Symptom experiences of midlife women: Observations from the Seattle Midlife Women's Health Study. *Maturitas*, *25*, 1–10.

Rusch, K. M., Guastello, S. J., & Mason, P. T. (1992). Differentiating symptom clusters of borderline personality disorder. *Journal of Clinical Psychology*, *48*, 730–738.

Shiffman, S. M., & Jarvik, M. E. (1976). Smoking withdrawal symptoms in two weeks of abstinence. *Psychopharmacology*, *50*, 35–39.

Siegel, J. P., Myers, B. J., & Dineen, M. K. (1987). Premenstrual tension syndrome symptom clusters: Statistical evaluation of the subsyndromes. *Journal of Reproductive Medicine*, *32*, 395–399.

Taub, E., Cuevas, J. L., Cook, E. W., 3rd, Crowell, M., & Whitehead, W. E. (1995). Irritable bowel syndrome defined by factor analysis. Gender and race comparisons. *Digestive Diseases and Sciences*, *40*, 2647–2655.

Westbrook, J. I., & Talley, N. J. (2002). Empiric clustering of dyspepsia into symptom subgroups: A population-based study. *Scandinavian Journal of Gastroenterology*, *37*, 917–923.

Woods, N. F., Mitchell, E. S., & Lentz, M. (1999). Premenstrual symptoms: Delineating symptom clusters. *Journal of Women's Health and Gender-Based Medicine*, *8*, 1053–1062.

Yardley, L., Masson, E., Verschuur, C., Haacke, N., & Luxon, L. (1992). Symptoms, anxiety and handicap in dizzy patients: Development of the vertigo symptom scale. *Journal of Psychosomatic Research*, *36*, 731–741.

Yasuda, N., Mino, Y., Koda, S., & Ohara, H. (2002). The differential influence of distinct clusters of psychiatric symptoms, as assessed by the general health questionnaire, on cause of death in older persons living in a rural community of Japan. *Journal of the American Geriatrics Society*, *50*, 313–320.